Which literature retrieval method is most effective for GPs?

Verhoeven, A.A.H.; Boerma, E.J.; Meyboom-de Jong, B.

Published in:
Family practice

Publication date:
2000

Family Practice
Vol. 17, No. 1

Printed in Great Britain

Which literature retrieval method is most effective for GPs?

Anita AH Verhoeven, Edzard J Boerma[a] and Betty Meyboom-de Jong[b]

Verhoeven AAH, Boerma EJ and Meyboom-de Jong B. Which literature retrieval method is most effective for GPs? *Family Practice* 2000; **17:** 30–35.

Background. Evidence-based medicine requires new skills of physicians, including literature searching.

Objective. To determine which literature retrieving method is most effective for GPs: the printed *Index Medicus*; Medline through Grateful Med; or Medline on CD-ROM.

Methods. The design was a randomized comparative study. In a continuing medical education course, three groups of health care professionals (87 GPs and 16 other health care professionals) used one of the literature retrieval methods to retrieve citations on four search topics related to general practice. For the analysis in pairs, we used the search results of the 75 participants who completed all four assignments. As outcome measures, we used precision, recall and an overall search quality score; we also had a post-course questionnaire on personal characteristics, experience with computers, handling medical literature and satisfaction with course instruction and search results.

Results. The recall and overall search quality scores in the *Index Medicus* groups ($n = 32$) were higher ($P = <0.001$) than those in the CD-ROM groups ($n = 31$). In addition, the search quality scores in the Grateful Med groups ($n = 12$) were higher ($P < 0.003$) than those in the CD-ROM groups. There were no differences in precision.

Conclusion. In the period 1994–1997, the printed *Index Medicus* was the most effective literature retrieval method for GPs. For inexperienced GPs, there is a need for training in electronic literature retrieval methods.

Keywords. Comparative study, family physicians, information storage and retrieval, MEDLARS, Medline.

Introduction

Due to the rapid expansion of medical knowledge and publications, physicians have difficulties in locating the medical information they need.[1] Experience, basic medical knowledge and skills alone are not enough to practise evidence-based medicine.[2] Physicians need to develop skills to retrieve and interpret information,[3] e.g. how to identify relevant diagnostic studies.[4] The medical literature is underused,[5] although 46–54% of primary care physicians' questions could be answered using the medical literature.[6,7]

In order to teach GPs the most effective literature retrieval method, we developed an experiment to compare three of these methods, namely a printed, an on-line and a CD-ROM version of the *Index Medicus*/Medline. The most effective method should give a high proportion of relevant citations of good quality, as well as few non-relevant citations. Furthermore, the ideal literature retrieval method should be widely available, easy to use, convenient, quick to learn and not expensive or time-consuming.[8]

In our experimental study, we answered the following research question: which literature retrieval method is most effective for GPs?

Received 8 April 1999; Revised 27 July 1999; Accepted 6 September 1999.
University Library, [a]Department of Education and [b]Department of General Practice, University of Groningen, Groningen, The Netherlands. Correspondence to Anita AH Verhoeven, University Library, University of Groningen, P.O. Box 559, 9700 Groningen, The Netherlands.

Methods

Participants

To invite GPs for the 1-day continuing medical education course 'How to retrieve information', we sent an announcement to all 970 registered GPs in the north of The Netherlands with reminders to university-affiliated GPs.

Randomization

Prior to sending the invitations, we determined the dates of the courses. We randomized the interested GPs to the *Index Medicus* course, the CD-ROM course or the Grateful Med course in order of receipt of the registration forms, and depending on the participants' day of choice. The participants did not know they would be randomized, because the invitation only mentioned one course including three methods in information retrieval.

Courses

From 1994 to 1997, we offered our 1-day course 15 times. A professional librarian, a GP by training, served as course instructor.

In each course, we focused on one of the following three literature retrieval methods: the 1992 printed *Cumulated Index Medicus*; Medline through Grateful Med version 6.0 (citations with publication year 1992) provided online by the Karolinska Institute in Sweden; and the 1992 Medline on CD-ROM (Silverplatter Dos version 3.1).

The experimental part of the course consisted of a 2-hour introduction in literature retrieval methods, and an on-site training session. In the introduction, the content and structure of one of the three methods were discussed, including controlled vocabulary, subheadings and, if appropriate, free-text searching. Sample searches were demonstrated.

The training session started with a try-out search by the participants in one of the three retrieval methods. Next, all three groups received the same four assignments for retrieving and selecting bibliographic citations. The four assignments covered search topics related to general practice: haemorrhoids, sudden infant death, the use of the telephone and the gatekeeper role (Table 1). For the Grateful Med and the CD-ROM groups, the order of the assignments was fixed and of increasing complexity. However, for organizational reasons, the assignments for the *Index Medicus* groups were performed in a varied order. The *Index Medicus* groups could spend 90 minutes in total on the four assignments. The Grateful Med and the CD-ROM group, however, could spend 80 minutes in total on the four assignments: 20 minutes for each assignment. Additionally, for these last two groups, the course instructor gave a 10-minute feedback directly after each assignment (Fig. 1).

After the experimental part of the course with one of the three retrieval methods, the two other methods were discussed and practised as well.

Questionnaire

After the course, the participants filled in a questionnaire which covered personal characteristics, experience with computers, handling medical literature and satisfaction with the course instruction and search results.

Outcome measures

The effectiveness of the searches was assessed by three measures: precision; recall; and an overall search quality score. Precision is the number of relevant citations as a proportion of the total number of citations retrieved; recall is the number of relevant citations retrieved from the total number of relevant citations in a subset of the bibliography (Table 2). In our study, this subset was formed out of all citations identified by the 103 course participants and the course instructor. Three judges assessed the relevance and the quality of these citations. Because GPs are more interested in quality than in numbers, we developed an overall search quality score. For this score, we calculated a citation quality score for each relevant citation, based on the following criteria: coverage of the journal by the Science or Social Science Citation Index or the Dutch list of Additional Scientific Journals for Health Sciences Research of the Royal Netherlands Academy of Arts and Sciences; the journal's impact factor; study design; and whether the citation was a review. To avoid negative values, we added 100 points to these scores. In a formula: the overall search

TABLE 1 *The four search queries for a course in literature searching for Dutch GPs*

I.	You have been invited to give a talk on haemorrhoids to a group of colleagues in your town. To collect information, you want to do a literature search using the *Index Medicus*/Medline on CD-ROM/Medline through Grateful Med for the year 1992.
II.	You want to keep up with the latest developments in the prevention of sudden infant death. You want to do a literature search using the *Index Medicus*/Medline on CD-ROM/Medline through Grateful Med for the year 1992.
III.	The editor of the Dutch journal *Huisarts & Wetenschap* (General Practitioner & Medical Science) has asked you to submit an article on the use of the telephone in the physician's office. You regard his request as a challenge and you decide to do a literature search using the *Index Medicus*/Medline on CD-ROM/Medline through Grateful Med for the year 1992.
IV.	You are supervising a GP trainee in your practice. She wants to discuss the gatekeeper role of the GP. You want to be well prepared for this discussion so you decide to do a literature search on this topic using the *Index Medicus*/Medline on CD-ROM/Medline through Grateful Med for the year 1992.

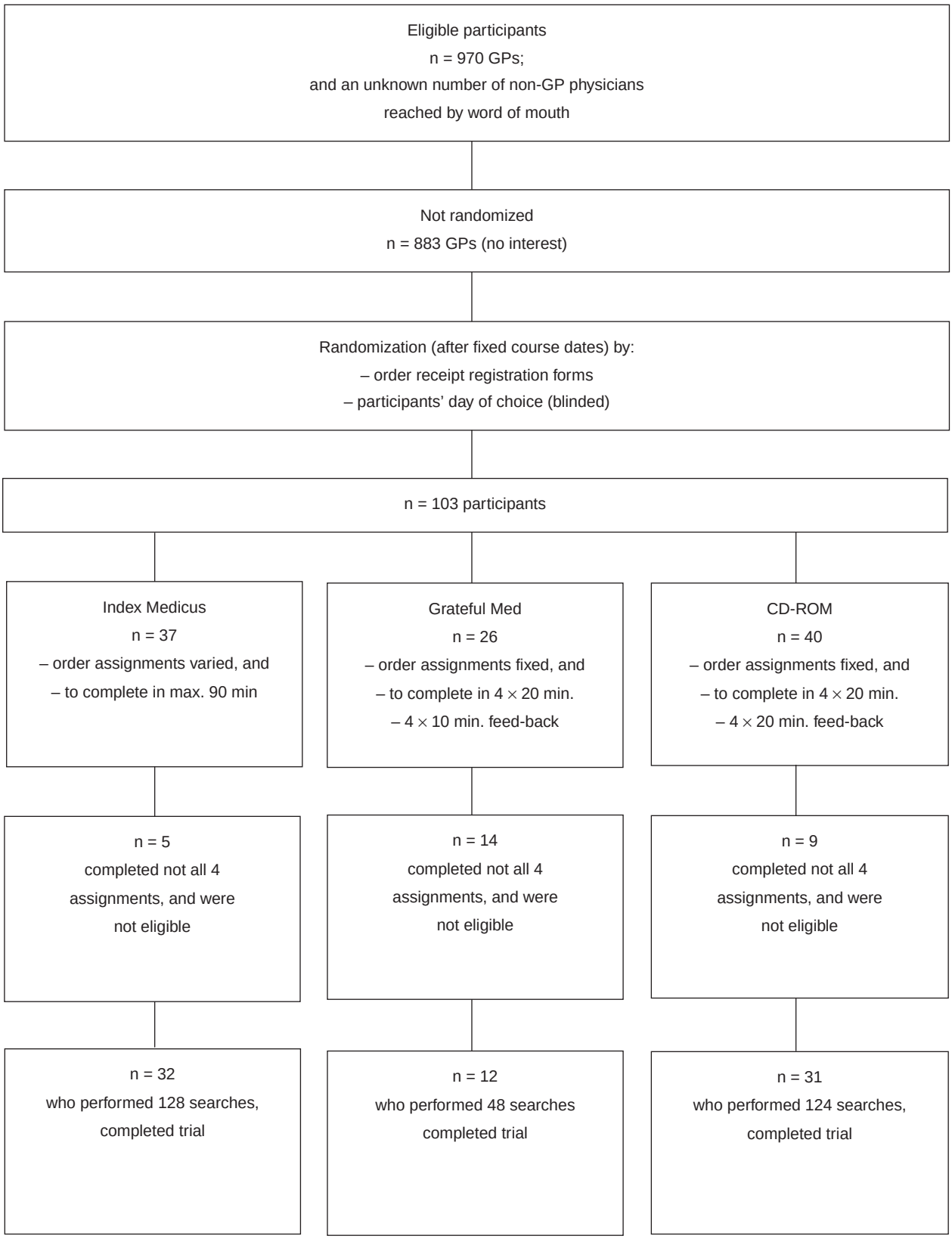

FIGURE 1 *Process through the various stages of randomization and intervention.*

TABLE 2 *Calculation of precision and recall of search results*

Assessed by course participants as	Assessed by the judges as	
	Relevant	Non-relevant
Relevant	*a*	*b*
Non-relevant	*c*	*d*

Precision = $a/(a + b)$; recall = $a/(a + c)$.

quality score was $a - b - c + 100$, in which a = total of quality scores of the selected relevant citations, b = total of quality scores of the missed relevant citations and c = number of the selected non-relevant citations.

Statistical analysis

We used SPSS version 8 for statistical analysis.

To check the equal allocation of the participants to the three groups, we compared sex, age, years of experience as a GP and type of practice with the chi-square and Student's *t*-tests.

To identify the influence of the method on precision, recall and overall search quality scores, a repeated measures analysis of variance (ANOVA) was used for each of the three methods (*Index Medicus*, Grateful Med and CD-ROM). The four assignments constituted the repeated factor in this set up; the method was the between-subjects factor.

To identify differences in precision, recall and search quality scores in pairs due to the method, in each of the four assignments, *post hoc* multiple comparisons for observed means were used according to Bonferroni.

To identify pairwise differences in precision, recall and overall search quality scores in each group for the four assignments in total, simple contrasts were used.

Results

Participants

Of the 970 invited GPs, 87 participants (9%) took part in the study. Additionally, 16 other health care professionals participated: seven nursing home physicians, six medical researchers in general practice, two physiotherapists and a hospital manager.

Of the 103 participants 73% were male. The mean age of the participants was 43 years. The GPs had 14 years of experience; 55% worked in an urban area; and 36% had a solo practice. Eighty-six per cent of all participants were affiliated to a University Department of General Practice.

Randomization

The 103 participants were randomized blockwise in 15 groups of 3–12 people (Fig. 1): 37 participants were randomized to the *Index Medicus* course (six groups), 26 to the Grateful Med course (four groups) and 40 to the CD-ROM course (five groups).

TABLE 3 *Precision, recall and search quality scores for 300 searches on four search topics in total performed using one of the three sources by 75 course participants*

Outcome measure	Literature retrieval method		
	Index Medicus $n = 32$	Grateful Med $n = 12$	CD-ROM $n = 31$
Precision	55	50	56
	(15)	(24)	(19)
Recall	26[a]	22	16
	(12)	(10)	(6)
Search quality score	83[b]	71[c]	65
	(5)	(6)	(6)

Results are means and (SD). n = participants who completed all four assignments.
[a]$P = 0.001$ versus CD-ROM; [b]$P < 0.001$ versus Grateful Med and CD-ROM; [c]$P = 0.003$ versus CD-ROM.

Outcome measures

The participants performed 366 searches: 310 by the 87 GPs, and 56 by the other health care professionals.

The allocation of the participants to the three groups was performed equally: the chi-square and *t*-tests showed no significant difference ($P < 0.01$) between the *Index Medicus*, Grateful Med and CD-ROM groups in sex or age, nor for the GPs in years of experience or type of practice. For the ANOVA and differences in pairs, we used the search results of the 75 participants (300 searches) who completed all four assignments. The repeated measures ANOVA showed a significant influence of the method on recall and search quality score ($P < 0.01$).

After combining the search results of the four search topics, the recall and overall search quality scores in the *Index Medicus* groups were significantly higher ($P < 0.001$) than the recall and overall search quality scores in the CD-ROM groups (Table 3). In addition, the search quality scores in the Grateful Med groups were higher ($P < 0.003$) than those in the CD-ROM groups. We found no difference in precision. The precision, recall and search quality scores of the individual search topics did not show specific patterns (Fig. 2).

Questionnaire

The questionnaire showed that 91% of the 103 participants possessed a personal computer for a mean 5.3 years: 88% used it for patient care, but only 14% for retrieving literature. Although 72% had written an article at least once, 45% had never visited an institutional library.

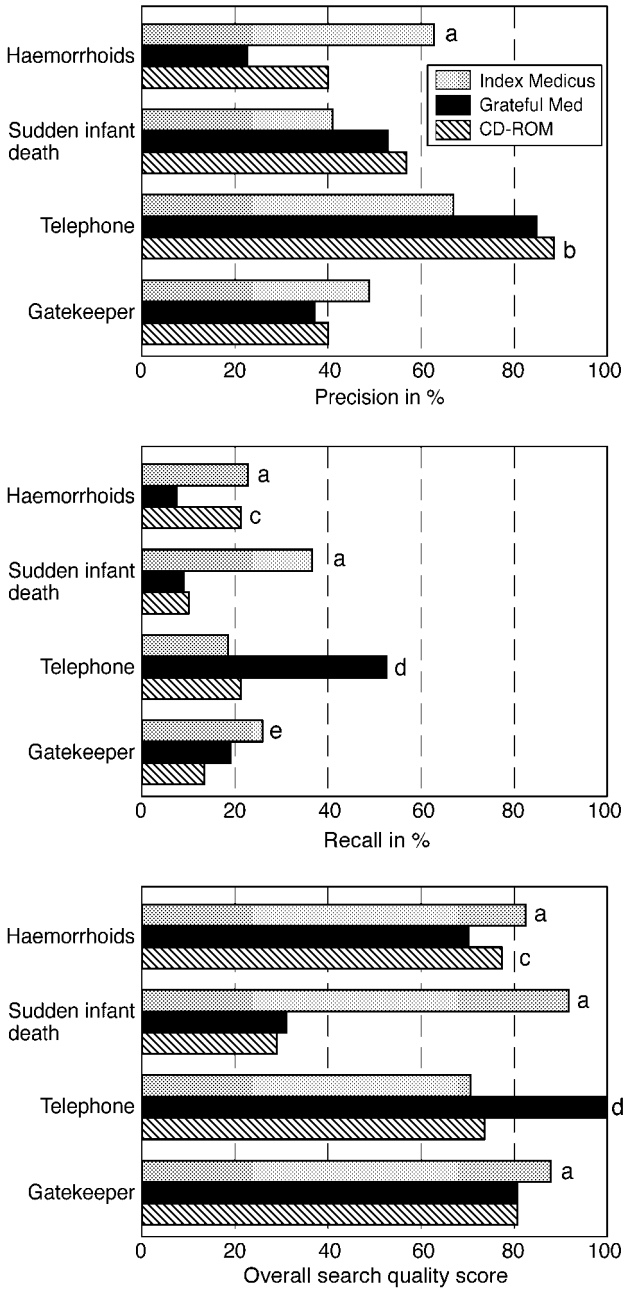

FIGURE 2 *Precision, recall and overall search quality scores for 300 searches on four individual search topics performed using one of the three sources by 75 course participants. Results are means. Precision is the proportion of citations retrieved by the course participant which were actually relevant. Recall is the proportion of all existing relevant citations which were actually retrieved by the course participant. The overall search quality score = a – b – c + 100, in which a = total of quality scores of the selected relevant citations, b = total of quality scores of the missed relevant citations and c = number of the selected non-relevant citations.*

[a]$P < 0.01$ versus CD-ROM, and versus Grateful Med; [b]$P < 0.01$ versus *Index Medicus*; [c]$P < 0.01$ versus Grateful Med; [d]$P < 0.01$ versus *Index Medicus*, and versus CD-ROM; [e]$P < 0.01$ versus CD-ROM

Almost all participants (99%) were satisfied with the course instruction, and 73% with the search results. Searching was interpreted as easy by 62%. The *Index Medicus*, though, was never used by 70% of the *Index Medicus* groups; Grateful Med was never used by 99% of the Grateful Med groups; and CD-ROM was never used by 77% of the CD-ROM groups.

Discussion

In this study performed from 1994 to 1997, we compared three methods for GPs with little experience of retrieving bibliographic information in literature searching. This study shows that the printed *Index Medicus* was the most effective literature retrieval method for GPs. Specifically, the printed *Index Medicus* yielded the best results in recall and overall search quality scores, whereas Medline on CD-ROM yielded the lowest scores. Apparently, the *Index Medicus* with only the Medical Subject Headings as entries was less confusing than Medline on CD-ROM which also offered free text searching.

No significant differences were found for precision. Obviously, the method used did not influence the critical selection of the retrieved citations. In fact, users consider precision of less importance than recall.[9]

New in this study was the use of the overall search quality score. Whereas the recall and precision referred to numbers of retrieved citations, the quality score took into account the quality of the relevant citations as also assessed by judges.

The results of our study could have been influenced by the following aspects of our study design.

(i) Since the course participants were interested volunteers, and 86% were affiliated to a University Department of General Practice, our participants could be more experienced in handling literature retrieval methods than GPs in general. Therefore, the results of our study cannot be generalized to the general population of GPs.

(ii) Although the *Index Medicus* group spent an average of 2.5 minutes longer on each assignment (10 minutes for all four assignments), it is unlikely that this influenced the results significantly.

(iii) In the CD-ROM and the Grateful Med groups, the assignment order was fixed and of increasing complexity. In addition, each assignment was discussed for 10 minutes afterwards. These factors could have had a positive influence on the learning effect, and therefore on the search results. In spite of the lack of this learning effect, it is remarkable that the *Index Medicus* group scored significantly better.

(iv) Although the topics of the assignments were related to daily practice, they were performed in a test situation with a restraint on time, and with a restriction of the information source (only 1 year). Actually,

the printed *Index Medicus* is only available in institutional libraries, whereas electronic bibliographic information sources can be accessed from home. Furthermore, perusing several years of the *Index Medicus* could be a time-consuming and tedious task.

(v) The test period was from 1994 to 1997. Because the literature retrieval software we used was current in the early 1990s, the study results are dated. Probably, with more consumer-friendly software packages, the same study performed in 1999 may show more favourable results for the electronic sources.

Although the printed *Index Medicus* yielded the best search results in our study, we would not recommend it to GPs. The *Index Medicus* may be the most effective literature retrieval method, but it does not seem to be the most efficient. GPs even prefer availability of information sources to quality.[10]

The implications of this study can be found in training programmes for GP trainees and continuing medical education courses. Because many search possibilities confused the inexperienced end users, training sessions need to be concise, clear and simple. A literature searching training session as part of an evidence-based medicine programme is recommended.

Because our study involved inexperienced end users with dated results, we expect that a repeated study with more experienced participants will give smaller or even reversed outcome measures of the three literature retrieval methods.

Conclusion

In summary, our study, carried out from 1994 to 1997, shows that out of the three, the printed *Index Medicus* is the most effective literature retrieval method for GPs. Because inexperienced end users had the highest recall and overall search quality scores using the simplest method, and the lowest recall and search quality scores using the method with the most search possibilities, we suggest the development of user-friendly electronic systems, and the provision of extensive training for end users.

Acknowledgements

We thank Henk Bosveld for his help with the statistical analysis.

References

[1] Huth EJ. The information explosion. *Bull NY Acad Med* 1989; **65:** 647–661.

[2] Guyatt GH. Evidence-based medicine: a new approach to teaching the practice of medicine. *J Am Med Assoc* 1992; **268:** 2420–2425.

[3] Greenhalgh T. *How to Read a Paper: The Basics of Evidence Based Medicine*. London: BMJ Publishing Group, 1997.

[4] van der Weijden T, IJzermans CJ, Dinant GJ, van Duijn NP, de Vet R, Buntinx F. Identifying relevant diagnostic studies in Medline. The diagnostic value of the erythrocyte sedimentation rate (ESR) and dipstick as an example. *Fam Pract* 1997; **14:** 204–208.

[5] Huth EJ. The underused medical literature. *Ann Intern Med* 1989; **110:** 99–100.

[6] Gorman PN, Ash J, Wykoff L. Can primary care physicians' questions be answered using the medical journal literature? *Bull Med Libr Assoc* 1994; **82:** 140–146.

[7] Chambliss ML, Conley J. Answering clinical questions. *J Fam Pract* 1996; **43:** 140–144.

[8] Smith R. What clinical information do doctors need? *Br Med J* 1996; **313:** 1062–1068.

[9] Su LT. The relevance of recall and precision in user evaluation. *J Am Soc Inform Sci* 1994; **45:** 207–217.

[10] Connelly DP, Rich EC, Curley SP, Kelly JT. Knowledge resource preferences of family physicians. *J Fam Pract* 1990; **30:** 353–359.